Advances in Applied Acoustics (AIAAS), Volume 5 2016 www.seipub.org/aiaas/
doi: 10.14355/aiaas.2016.05.001

A PLCA Model for Detection of Humpback Whale Sound Units

Dorian Cazau*[1], Olivier Adam[2]

[1]ENSTA Bretagne - Lab-STICC (UMR CNRS 6285), Université Européenne de Bretagne, 2 rue François Verny, 29806 Brest Cedex 09, France

[2]Sorbonne Universités, UPMC Univs Paris 6, UMR 7190, Institut JeanLe Rond d'Alembert 4, Place Jussieu, 75252 PARIS Cedex 05

*cazaudorian@outlook.fr

Abstract

In this paper, we explore the use of the Probabilistic Latent Component Analysis (PLCA) method to detect sound units of real humpback whale songs. Through a dictionary of vector templates, an acoustic signature can be explicitly assigned to each sound event to be retrieved, while allowing the incorporation of prior parameter to constrain template definition. This method provides a powerful framework for modeling and retrieving of audio information in complex sound mixtures, and this first application to humpback whale songs has already shown great promise.

Keywords

PLCA; Machine Learning; Humpback Whale Songs; Bioacoustics

Introduction

Scientists have been interested in the complex stereotyped songs that male individuals of humpback whales (*Megaptera novaeangliae*) emit during the winter-spring breeding season. Such a complex acoustic display is thought to play an important role in both the mating ritual and male to male interaction (see e.g. [1]). Hence, the need to detect and further classify the unit constituents of a song objectively and systematically has become crucial to analyse the songs and to allow processing large data set. In this paper, we propose to task these difficulties using the Probabilistic Latent Component Analysis (PLCA) method. PLCA belongs to a class of probabilistic models, known as latent class models, that attempt to explain the observed histograms as having been drawn from a set of latent classes, each with its own distribution. PLCA has been developed as a general method for feature extraction from non-negative data with pioneering applications to audio [2]. This paper presents for the first time an original application of PLCA to the tasks of detection of sound units of humpback whale songs. Past computational methods for these tasks (e.g. sparse coding approach in [3] and blind clustering approach in [4]) were limited in terms of modeling flexibility and result interpretation with detection performance degrading in noise. Our approach is semi-supervised, using both pre-trained models as in [5], and uses a template adaptation process to better our spectral basis to the data. Then, by explicitly modeling both underwater noise and vocal events in humpback whale songs, and allowing interpretative analysis of these basis, PLCA should provide a powerful framework for the representation and understanding of their structures. We first present in Sec. II-A the theoretical background of PLCA, then propose an adaptation of this framework for detection of humpback whale sound units in Sec. II-B, and eventually provide in Sec. III preliminary results and discussion on this original application of PLCA to whale sounds.

Methods

Probability Latent Component Analysis (PLCA)

A suitably normalized magnitude spectrogram $X(f,t)$ can be modeled as a joint distribution over time and frequency, $P(f,t)$, with f and t respectively the frequency and time indexes. We can decompose it as a product of a spectral basis matrix and component activity matrix, as follows

$$P(f,t) = P(t)\sum_{z} P(f|z)P(z|t) \quad (1)$$

where z is the component index, $P(t)$ is the energy of the input spectrogram (known quantity), $P(f|z)$ is the spectral template that corresponds to the z^{th} component, and $P(z|t)$ is the activation of the z^{th} component. To estimate the model parameters $P(f|z)$ and $P(z|t)$, since there is usually no closed-form solution for the maximization of the loglikelihood or the posterior distributions, iterative update rules based on the Expectation-Maximization (EM) algorithm [6] are classically employed (see [2] for details).

Two-Class PLCA for Sound/Noise Detection

The frequency marginals $P(f|z)$ can be used as a model for certain kinds of sounds, which will now allow us to tackle the tasks of detection using separate class modeling. In a given humpback whale song, two sound classes are assumed to be present, namely the background noise from the marine environment (composed of sound pollution from boat traffic, rain and wind at the sea surface, water current, animal movements, sometimes even bubbles and rockfall …), and the vocal sound units produced by humpback whales.

In the following, we then define two different template dictionaries, whose elements are defined by the frequency marginals distribution $P_v(f|z)$ and $P_n(f|z)$, with $z \in Z_v$ and Z_n, respectively the sets of latent variables for the vocalization and noise classes.

1) Knowledge-Based Initialization of Template Dictionaries

A crucial step before performing a PLCA-based analysis is to properly initialize template dictionaries, which allows EM-based parameter estimation to be much more precise [7]. To do so, we use knowledge from bioacoustics literature on humpback whale songs. We adopt the results from [8, Fig. 2], who defined 9 different stereotypical call types distinguished from each other by salient acoustic properties (e.g. rich harmonic sounds with high frequency for unit 1, and broadband noise like sound for unit 2). We then initialize our different $P_v(f|z)$, performing successively 9 one-component PLCA for each of these 9 vocal sound units. Similarly, we also initialize $P_n(f|z)$ using 3 stereotypical underwater noise sounds, extracted from our hydrophone recordings, and identified to the sources of boat traffic, rain at the sea surface and whale movements. The number of latent variables for the noise class has been arbitrarily set after exploring the diversity of a large set of noise sounds.

2) PLCA for Conservative Detection

The concept of conservative segmentation [9] consists of identifying only those detected events for which we have a high degree of confidence, and omitting any unsure candidates. Such events are likely to fit well the main acoustic constituents in our recordings. This conservative segmentation is performed with a PLCA model presenting two sets of frequency marginals, and is defined as follows

$$\begin{aligned} X(f,t) \approx P(f,t) = P(t)\Big(&\sum_{z\in Z_n} P(f|z)P(z|t) \\ &+ \sum_{z\in Z_v} P(f|z)P(z|t)\Big) \end{aligned} \quad (2)$$

where we used fixed template dictionaries defined in the previous section. Note events are then extracted by thresholding the event activation matrix $P(z,t) = P(t)P(z|t)$, and select only time intervals whose activity values exceed the threshold α. This parameter controls the levels of precision/recall: a low threshold has a high recall and low precision; the opposite occurs with a high threshold, which is done for our conservative segmentation. Eventually, two collections of time frames $\boldsymbol{T}_v = t_1,\ldots, t_N$, and a similar one $\boldsymbol{T}_n$, are obtained, which can be used in order to adapt our template dictionaries to data.

3) Data-Based Learning of Template Dictionaries

The previous processing allowed us to identify the time intervals of highly probable occurrences of vocal sounds and background noise, respectively $\boldsymbol{T}_v$ and $\boldsymbol{T}_n$. We will now use these intervals to learn the two template dictionaries corresponding to the background noise and the whale vocal sounds, adaptively to the input data. As mentioned in Sec. II-B1, a one-component PLCA is used for each time interval, taking as input $S(f,t_i)$ with

$t_i \in T_k$, with $k \in \{v, n\}$ the indexes respectively for vocalization and noise classes. The output for each latent component z_k is a spectral template $P(f|z_k)$ which can be used in order to expand the present dictionary. As in [10], this template adaptation can be controled by a parameter ι through the following equation

$$P(f,t) = \frac{\sum_t \iota_k P_t(z_k|f) X(f,t) + (1-\iota_k)\omega_{theo}}{\sum_{f,t} \iota P_t(z_k|f) X(f,t) + (1-\iota_k)\omega_{theo}} \quad (3)$$

with $w_{theo} = P(f|z_k)$ corresponding to the template dictionaries learned in Sec. II-B1. $P_t(z|f)$ is the posterior of the model (defined in [11]). Setting higher values of the parameter ι allows increasing template adaptation.

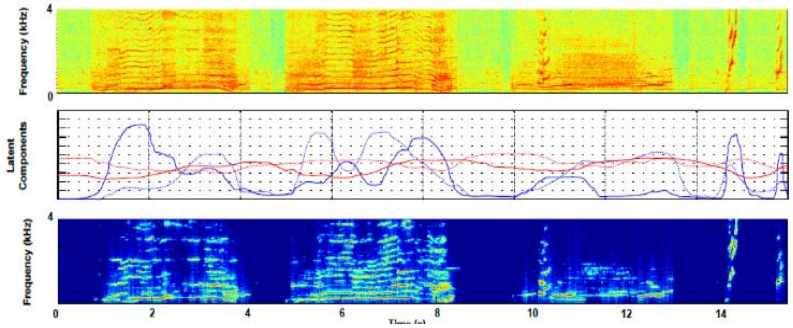

FIG. 1 EXAMPLE OF AN ANALYZED SEQUENCE, SHOWING FROM TOP TO BOTTOM: SPECTROGRAM, TIME ACTIVATIONS OF FOUR DIFFERENT LATENT VARIABLES (TWO FROM THE VOCAL CLASS, IN BLUE, AND TWO FROM THE NOISE CLASS, IN RED) AND THE PLCA RECONSTRUCTED SPECTROGRAM USING ONLY THE LATENT VARIABLES OF THE VOCAL CLASS.

4) Detection of Vocal Sounds

Here, we still use the PLCA model developed in eq. 4, but now, we only update templates corresponding to the components $z \in Z_v$, as we assume that the background noise is locally much less variable than humpback whale sound units. Our detection process starts with extracting the vocalization events from the mixture as follows. Assuming that the background noise stays stable during the entire duration of a whale song, we can set noise activations $P(z)$ for $z \in Z_n$ to zero, and therefore, a denoised reconstructed PLCA model of sound spectra $P_v(f,t)$ can be obtained by:

$$P_v(f,t) \approx P(t) \sum_{z \in Z_v} P(f|z) P(z|t) \quad (4)$$

After this denoising process, we use a simple threshold-based detection of the vocal event activations from the activity matrix $P(z,t)$, followed by a minimum duration pruning. The threshold of minimum duration for pruning was set to 200 ms. An analyzed sequence is shown in figure 1.

Experiments

Evaluation Procedure

4 Humpback whales songs were used for evaluation, and two hours of audio with hand-labeled sound units were used for evaluation. These songs were recorded in the Sainte Marie Island Channel, North East Madagascar, during Aug. 2007, Aug.2008 and Aug.2009. Recordings were done from the boat (motor off) using the ColmarItalia GP280 hydrophone and digitalized by the Tascam HD-P2 recorder at a sample frequency (Fs) of 44.1 kHz and coded on 16 bits. For assessing the performance of our proposed detection system, we adopt an event-oriented approach, according to which a vocal event is assumed to be correct if it fulfills the condition that its onset is within a 0.5 sec range from a ground-truth onset, and that its duration is within 20% of the ground truth note duration. Evaluation metrics are defined by the event-based recall (TPR), precision (PPV) and F (the F-measure, i.e. the harmonic mean of precision and recall).

Results and Discussion

Table I compares detection performance of our method with two other systems: PamGuard [12] and Ishmael [13].

Both of these systems are publically available softwares[1], and have been widely used in the marine mammal community for the task of vocal sound detection in particular. From table I, it is obvious that the proposed algorithm performs better than other systems, with a F value of 69.7 %, in comparison to 65.2 % 61.6 % for PamGuard and Ishmael, respectively. Furthermore, when the vocal sounds are submerged in noise entirely, the signal can be extracted out well and detected after processing the received signal by the proposed algorithm. False alarms are in particular at the lowest for our system, which proves that a better discrimination between vocal and noise sounds is obtained through our two-class PLCA model. Figure 2 presents the evolution of a relative metric ΔF - *measure* = F - $F_{Default}$, against the parameters ι_V and ι_N. $F_{Default}$ is the the F - *measure* obtained with the default values of ι_V and ι_N, both set to the value of 0.05.

TABLE I ERROR METRICS OBTAINED WITH THE DIFFERENT DETECTION SYSTEMS, AND AVERAGED OVER THE FOUR DIFFERENT SONGS.

Methods	TPR	PPV	F
Pamguard	63.2%	67.4%	65.2%
Ishmael	60.1%	63.3%	61.6%
Proposed method	**68.9%**	**70.5%**	**69.7%**

We can see that these two parameters impact differently the four humpback whale songs. Higher values of the ι means a more important contribution of the data-based template adaptation to detection performance. One can observe from these results that our two longest humpback whale songs, namely S2 and S4, benefit the most from higher values of ι parameters. We could explain this tendency by the fact that longer songs encompass a huger acoustic diversity, both in vocal sounds and environmental background noise, which should deviate more from our a priori knowledge-based templates (defined in Sec. II-B1), and thus need to be more adapted to input data.

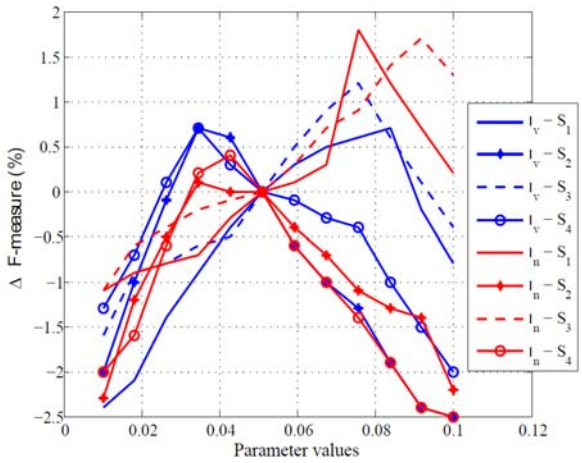

FIG. 2 EVOLUTION OF THE F METRIC AGAINST THE PARAMETERS I_V AND I_N, FOR THE DIFFERENT SONGS S_1 TO S_4.

Conclusions

In this paper, we have developed the use of the Probabilistic Latent Component Analysis (PLCA) method to detect sound units of real humpback whale songs. In contrast to blind clustering method, our approach combines a supervised learning of our template dictionaries using a priori bioacoustics knowledge, and also a semi-supervised procedure to adapt these templates to input data. Our PLCA-based method for detection could indeed constitute a powerful tool of wildlife assessment and provide crucial information to understand how whales are affected by disturbance and noise pollution from energy exploration, shipping, and other human activities, and to advise industry and government on how to minimize harm to marine wildlife.

ACKNOWLEDGMENT

This work was partially financially supported by the Association Dirac (France).

[1] Downloading links of these softwares are http://www.bioacoustics.us/ishmael.html and http://www.pamguard.org/

REFERENCES

[1] P. Tyack, “Interactions between singing hawaiian humpback whales and conspecifics nearby,” Behavioral Ecology and Sociobiology, vol. 8, pp. 105–116, 1981.

[2] P. Smaragdis, B. Raj, and M. Shanshanka, “A probabilistic latent variable model for acoustic modeling,” in Neural Information Proc. Systems Workshop, Whistler, BC, Canada, 2006.

[3] X. C. Halkias, S. Paris, and H. Glotin, “Classification of mysticete sounds using machine learning techniques,” J. Acoust. Soc. Am., vol. 134, pp. 3496–3505, 2013.

[4] H. Ou, W. W. L. Au, L. M. Zurk, and M. O. Lammers, “Automated extraction and classification of time-frequency contours in humpback vocalizations,” J. Acoust. Soc. Am., vol. 133, pp. 301–310, 2013.

[5] G. Grindlay and D. P. W. Ellis, “A probabilistic subspace model for multi-instrument polyphonic transcripton,” in 11th InternationalSociety for Music Information Retrieval Conference, Utrecht, Netherlands, 2010.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” Journal of the Royal Statistical Society, Series B, vol. 39, pp. 1–38, 1977.

[7] T. Cheng, S. Dixon, and M. Mauch, “A deterministic annealing algorithm for automatic music transcription,” in 14th International Society for Music Information Retrieval Conference, Curitiba, PR, Brazil, 2013.

[8] W. W. L. Au, A. A. Pack, M. O. Lammers, L. M. Herman, M. H. Deakos, and K. Andrews, “Acoustic properties of humpback whale songs,” J. Acoust. Soc. Am., vol. 120, pp. 1103–1110, 2006.

[9] D. Tidhar, M. Mauch, and S. Dixon, “High precision frequency estimation for harpsichord tuning classification.” in IEEE Int. Conf. Audio, Speech and Signal Processing, Dallas, USA, 2010.

[10] E. Benetos, R. Badeau, T. Weyde, and G. Richard, “Template adaptation for improving automatic music transcription,” in 15th International Society for Music Information Retrieval Conference, Taipei, Taiwan, 2014, pp. 175–180.

[11] E. Benetos, S. Cherla, and T. Weyde, “An efficient shift-invariant model for polyphonic music transcription,” in 6th Int. Workshop on Machine Learning and Music, Prague, Czech Republic, 2013.

[12] D. Gillespie, “Pamguard: semiautomated, open source software for realtime acoustic detection and localization of cetaceans,” in Proceedings of the Institute of Acoustics, 2008.

[13] D. Mellinger, “Ishmael 1.0 users guide,” Natl. Oceanogr. Atmos. Admin. Tech. Memo. OAR-PMEL-120 (NOAA PMEL, Seattle), 30 pp, Tech. Rep., 2001.

Dorian Cazau He received the B. Sc. (Fundamental Physics) and M. Sc. (Acoustics and Signal Processing) from the University Pierre and Marie Curie (Paris, France) in 2010 and 2012. Currently, he is working as a post-doctoral researcher at the LabSTICC (ENSTA-Bretagne, Brest, France). His research interests include audio information retrieval, acoustic space modeling, with applications to different complex acoustic sciences , from Bioacoustics to Music Information Retrieval.

Olivier Adam Full Professor at the University Pierre et Marie Curie (Paris, France). Specialist in Signal Processing, he works on bioacoustics since 2002, especially on sounds emitted by cetacean species. He is currently involved in a study on the sound generator of baleen whales, in collaboration with Dr. Joy Reidenberg (New York). In 2010, he came to St Pierre-et-Miquelon for his first time and was engaged on the acoustic observations (materials, deployment, analysis).